

# Quantum Information on Spectral sets

Peter Harremoës

Niels Brock

Copenhagen Business College

Copenhagen

E-mail: harremoes@iee.org

**Abstract**—For convex optimization problems Bregman divergences appear as regret functions. Such regret functions can be defined on any convex set but if a sufficiency condition is added the regret function must be proportional to information divergence and the convex set must be spectral. Spectral sets are sets where different orthogonal decompositions of a state into pure states have unique mixing coefficients. Only on such spectral sets it is possible to define well behaved information theoretic quantities like entropy and divergence. It is only possible to perform measurements in a reversible way if the state space is spectral. The most important spectral sets can be represented as positive elements of Jordan algebras with trace 1. This means that Jordan algebras provide a natural framework for studying quantum information. We compare information theory on Hilbert spaces with information theory in more general Jordan algebras, and conclude that much of the formalism is unchanged but also identify some important differences.

## I. INTRODUCTION

Although quantum physics has existed for more than 100 years there are still many simple fundamental questions that remain unanswered. One of the main questions is why complex Hilbert spaces are used. Although the complex numbers are extremely useful for doing computations we do not have direct observations of these numbers. Observables are represented by selfadjoint operators and such operators can be multiplied but the multiplication leads to operators that are not selfadjoint and cannot be observed. Therefore Pascual Jordan introduced what is now known as Jordan algebras with a different product than the usual matrix product. Still, his product is not well physically motivated. In this paper we give some general definitions of some information theoretic quantities and demonstrate that they are only well behaved on convex sets that can be represented on Jordan algebras. Therefore the Jordan algebras appear to be the correct formalism for doing quantum information. From a computer science point of view the quantum algorithms should be run in Jordan algebras, but for building a quantum computer we need a physical implementation of the algorithm. Apparently the physical world prefer Jordan algebras associated with complex Hilbert spaces so a general algorithm on a Jordan algebra should be implemented on complex Jordan algebras. This is possible in most cases but may increase the number of gate operations by a factor.

## II. STRUCTURE OF THE STATE SPACE

Our knowledge about a system will be represented by a state space. In many cases the state space is given by a set of

probability distributions on the sample space. In such cases the state space is a simplex, but it is well-known that the state space is not a simplex in quantum physics. In order to cover applications in physics we need a more general notion of a state space as defined in [1].

### A. The state space and the positive cone

Before we do any measurement we prepare our system. Let  $\mathcal{P}$  denote the set of preparations. Let  $p_0$  and  $p_1$  denote two preparations. For  $t \in [0, 1]$  we define  $(1 - t) \cdot p_0 + t \cdot p_1$  as the preparation obtained by preparing  $p_0$  with probability  $1 - t$  and  $p_1$  with probability  $t$ . A measurement  $m$  is defined as an affine mapping of the set of preparations into a set of probability measures on some measurable space. Let  $\mathcal{M}$  denote a set of feasible measurements. The state space  $\mathcal{S}$  is defined as the set of preparations modulo measurements. Thus, if  $p_1$  and  $p_2$  are preparations then they represent the same state if  $m(p_1) = m(p_2)$  for any  $m \in \mathcal{M}$ .

In statistics the state space equals the set of preparations and has the shape of a simplex. The symmetry group of a simplex is simply the group of permutations of the extreme points. In quantum physics the state space has the shape of the density matrices on a complex Hilbert space and the state space has a lot of symmetries that a finite simplex does not have. For simplicity we will assume that the state space is a finite dimensional convex compact space.

From now on we consider the situation where our knowledge about a system is given by an element in a convex set. These elements are called *states* and convex combinations are formed by probabilistic mixing. States that cannot be distinguished by any measurement are considered as being the same state. The extreme points in the convex set are called *pure states* and all other states are called *mixed states*. See [1] for details about this definition of a state space.

If  $\mathcal{S}$  is a state space it is sometimes convenient to consider the positive cone generated by  $\mathcal{S}$ . The positive cone consists of elements of the form  $\lambda \cdot \sigma$  where  $\lambda \geq 0$  and  $\sigma \in \mathcal{S}$ . Elements of a positive cone can be multiplied by positive constants via  $\lambda \cdot (\mu \cdot \sigma) = (\lambda \cdot \mu) \cdot \sigma$  and can be added as follows.

$$\lambda \cdot \rho + \mu \cdot \sigma = (\lambda + \mu) \cdot \left( \frac{\lambda}{\lambda + \mu} \rho + \frac{\mu}{\lambda + \mu} \sigma \right).$$

The convex set and the positive cone can be embedded in a real vector space by taking the affine hull of the cone and use the apex of the cone as origin of the vector space.

### B. Measurements on the state space

Let  $m$  denote a measurement on the state space  $S$  with values in the set  $A$ . Then  $m(\sigma)$  is a probability distribution on  $A$ . Let  $B \subseteq A$ . Then  $m(\sigma)(B)$  is the probability of observing a result in  $B$  when the state is  $\sigma$  and the measurement  $m$  is performed. Then  $m(\sigma)(B) \in [0, 1]$  and  $\sigma \rightarrow m(\sigma)(B)$  is an affine mapping.

**Definition 1.** Let  $S$  denote a convex set. A *test* is an affine map from  $S$  to  $[0, 1]$ .

The tests are building block for all measurements according to the following proposition.

**Proposition 2.** Let  $m$  denote a measurement on the convex set  $S$  with values in the finite set  $A$ . Then there exists tests  $\phi_b : S \rightarrow [0, 1]$  such that for any  $B \subseteq A$  we have

$$m(\sigma)(B) = \sum_{b \in B} \phi_b(\sigma). \quad (1)$$

If the set  $A$  is not finite we may have to replace the sum by an integral. The trace of a positive element is defined by  $\text{tr}(\lambda \cdot \sigma) = \lambda$  when  $\sigma \in S$  so that states are positive elements with trace equal to 1. We note that the trace restricted to the states defines a test. Any tests can be identified with a positive functional on the positive cone that is dominated by the trace. Note that if a measurement  $m$  is given by (1) then the tests satisfies

$$\sum_{b \in A} \phi_b = \text{tr}.$$

### C. Improved Caratheodory theorem

Let  $x$  be an element in the positive cone such that

$$x = \sum_{i=1}^n \lambda_i \cdot \sigma_i. \quad (2)$$

where  $\sigma_i$  are pure states. If  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  then the vector  $\lambda_1^n$  is called the *spectrum of the decomposition*. Note that there are no restrictions on the number  $n$  in the definition of the spectrum, so if two spectra have different length we will extend the shorter vector by concatenating zeros at the end. We note that for a decomposition like (2) the trace is given by  $\text{tr}[x] = \sum_{i=1}^n \lambda_i$ .

Spectra are ordered by *majorization*. Let  $\lambda_1^n$  and  $\mu_1^n$  be the spectra of two decompositions of the same positive element. Then  $\lambda_1^n \succeq \mu_1^n$  if  $\sum_{i=1}^k \lambda_i \geq \sum_{i=1}^k \mu_i$  for  $k \leq n$ . Note that in a general positive cone the majorization ordering is a partial ordering. In special cases like the cone of positive semidefinite matrices on a complex Hilbert space the decompositions of the matrix form a lattice ordering with a unique maximal element, but in general the set of decompositions may have several incompatible maximal elements.

**Definition 3.** Let  $S$  denote a convex set. Let  $\sigma_i, i = 1, 2, \dots, n$  states in the state space  $S$ . Then  $(\sigma_i)_i$  are said to be perfectly distinguishable if there exists tests  $\phi_i$  such that  $\phi_i(\sigma_j) = \delta_{ij}$ . The states  $\sigma_0$  and  $\sigma_1$  are said to be *orthogonal* if  $\sigma_0$  and  $\sigma_1$  are perfectly distinguishable in the smallest face  $F$  of

$S$  that contain both  $\sigma_0$  and  $\sigma_1$ . If the extreme points  $\sigma_1, \sigma_2, \dots, \sigma_n$  of a decomposition are orthogonal then the decomposition is called an *orthogonal decomposition*.

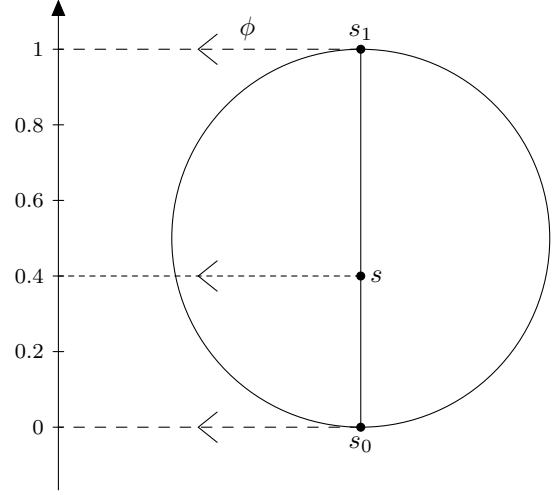


Fig. 1. In the disc the points  $\sigma_0$  and  $\sigma_1$  are mutually singular. The point  $s$  has a unique decomposition into mutually singular points because it is not the center of the disc.

**Theorem 4.** Let  $S$  denote a convex compact set of dimension  $d$  and let  $x$  denote some element in the positive cone generated by  $S$ . Then there exists an orthogonal decomposition of the form as in Equation 2 such that  $n \leq d + 1$ .

*Proof:* See [2, Thm. 2]. ■

**Definition 5.** The *rank* of a convex set is the maximal number of orthogonal states needed in an orthogonal decomposition of a state.

**Example 6.** In the unit square with  $(0, 0)$ ,  $(1, 0)$ ,  $(0, 1)$  and  $(1, 1)$  as vertices the point, with coordinates  $(1/2, 1/4)$  has an orthogonal decomposition with spectrum  $(1/2, 1/4, 1/4)$ . This spectrum majorizes the spectrum of any other decomposition of this point, and it also majorizes the spectrum of any other point in the square. The square has in total four points symmetrically arranged with the same spectrum as  $(1/2, 1/4)$ .

### III. SPECTRAL SETS

Any state in may be decomposed into orthogonal states, but such orthogonal decompositions are only unique when the state space is a simplex. Nevertheless there exists a type of convex sets where some weaker type uniqueness holds.

#### A. Spectrality conditions

**Definition 7** ([3]). An  $n$ -frame is a list of  $n$  perfectly distinguishable pure states. If any state  $\sigma \in S$  has a decomposition  $\sigma = \sum_i \lambda_i \cdot \sigma_i$  where  $\sigma_1, \sigma_2, \dots, \sigma_n$  is an  $n$ -frame then the state space is said to satisfy weak spectrality.

**Proposition 8.** Let  $S$  denote a state space such that

1. If both  $\rho \perp \sigma_1$  and  $\rho \perp \sigma_2$  then  $\rho \perp \frac{1}{2}(\sigma_1 + \sigma_2)$ .
2. If  $\sigma_1 \perp \sigma_2$  then  $\sigma_1$  and  $\sigma_2$  are perfectly distinguishable.

Then each state satisfies weak spectrality.

**Definition 9.** If all orthogonal decompositions of a state have the same spectrum then the common spectrum is called it the *spectrum of the state* and the state is said to be *spectral*. We say that a state space  $\mathcal{S}$  is *spectral* if all states in  $\mathcal{S}$  are spectral.

Often we are interested in a stronger condition.

**Definition 10.** Let  $\sigma$  denote a state with the orthogonal decompositions  $\sigma = \sum s_i \sigma_i = \sum r_j \rho_j$ . We say that the decompositions are strictly spectral if  $\sum f(s_i) \sigma_i = \sum f(r_j) \rho_j$  for any real valued function  $f$ . A state is strictly spectral if any orthogonal decomposition is strictly spectral. A state space is strictly spectral if any state is strictly spectral.

The important case is when the function  $f$  equals  $1_\lambda$  for which we get that  $\sum s_i \sigma_i = \sum r_j \rho_j$  implies that  $\sum_{s_i=\lambda} \sigma_i = \sum_{r_j=\lambda} \rho_j$ . An element of the form  $\sum_{s_i=\lambda} \sigma_i$  is called an idempotent. Strict spectrality means that a state has a unique decomposition  $\sigma = \sum \lambda_k e_k$  where  $\lambda_k$  are distinct eigenvalues and  $e_k$  are orthogonal idempotents. We note that if the state space is strictly spectral then for any element  $x$  in the generated real vector space we have  $y \geq 0$  if and only if there exists  $x$  such that  $y = x^2$ . In particular  $\sum x_i^2 = 0$  if and only if  $x_i = 0$  for all  $i$ .

### B. Jordan algebras

The notion of a spectral set is related to self-duality of the cone of positive elements, which leads to the following conjecture.

**Conjecture 11.** If a finite dimensional convex compact set is spectral and has a transitive symmetry group then the convex set can be represented as positive elements with trace 1 in a simple Jordan algebra.

The density matrices with complex entries play a crucial role in the mathematical theory of quantum mechanics and it is well-known that the density matrices is a spectral set. For each density matrix the spectrum equals the usual spectrum calculated as roots of the characteristic polynomial. In the 1930'ties P. Jordan generalized the notion of Hermitean complex matrix to the notion of a Jordan algebra in an attempt to provide an alternative to the complex Hilbert spaces as the mathematical basis of quantum mechanics. For instance the complex Hermitean matrices form a Jordan algebra with the *quasi-product* defined by

$$x \circ y = \frac{1}{2} \left( (x+y)^2 - x^2 - y^2 \right). \quad (3)$$

A direct expansion of the squares lead to  $x \circ y = \frac{1}{2} (xy + yx)$ .

An formally real Jordan algebra is an algebra with composition  $\circ$  that is commutative and satisfies the Jordan identity

$$(x \circ y) \circ (x \circ x) = x \circ (y \circ (x \circ x)).$$

Further it is assumed that

$$\sum_{i=1}^n x_i^2 = 0$$

implies that  $x_i = 0$  for all  $i$ . In an formally real Jordan algebra we write  $x \geq 0$  if  $x$  is a sum of squares.

For a matrix  $(M_{mn})$  the trace  $\text{tr}$  is defined by  $\text{tr}[M] = \sum_n M_{nn}$ . The trace vanish on commutators and associators. For an associative algebra it means that  $\text{tr}[MN] = \text{tr}[NM]$  and for a Jordan algebra it means that  $\text{tr}[(a \circ b) \circ c] = \text{tr}[a \circ (b \circ c)]$  (see [4] for details). For a Jordan algebra one can define an inner product by  $\langle x, y \rangle = \text{tr}[x \circ y]$ . With this inner product the positive cone becomes self-dual.

In any finite dimensional formally real Jordan algebra we may define the density operators as the positive elements with trace 1. Then the density operators of a Jordan algebra is a spectral set. Any formally real Jordan algebra is a sum of simple formally real Jordan algebras. There are 5 types of simple Euclidian Jordan algebras leading to the following convex sets:

**Spin factor** A unit ball in  $d$  real dimensions.

**Real**  $n \times n$  density matrices over the real numbers.

**Complex**  $n \times n$  density matrices over complex numbers

**Quaternionic**  $n \times n$  density matrices over quaternionians.

**Albert**  $3 \times 3$  density matrices with octonian entries.

The first four types are called the *special types* and the last one is called the *exceptional type*. In each of the four special Jordan algebras the product is defined by (3) from an associative product. This is not possible for the Albert algebra, which is the reason that it is said to be exceptional. See [5] for general results on Jordan algebras and [6] for details about quantum mechanics based on quaternions. The  $2 \times 2$  matrices have real, complex, quaternionic or octonionic entries can be identified with spin factors with  $d = 2, d = 3, d = 5$ , and  $d = 9$ . The most important example of a spin factor is the qubit.

A sum of different simple Jordan algebras does not fulfill the strong symmetry mentioned in [3] although it is spectral and fulfill projectivity. This was left as an open question in [3].

### C. Separable states

Let  $\mathbb{H}_1 \otimes \mathbb{H}_2$  denote a tensor product of two complex Hilbert spaces. Then  $\mathbb{B}(\mathbb{H}_1 \otimes \mathbb{H}_2) = \mathbb{B}(\mathbb{H}_1) \otimes \mathbb{B}(\mathbb{H}_2)$ . The separable states are mixtures of states of the form  $\sigma_1 \otimes \sigma_2$  where  $\sigma_i$  is a density operator in  $\mathbb{B}(\mathbb{H}_i)$ . We will show that the set of separable states is spectral.

First we note that for  $A \in \mathbb{B}(\mathbb{H}_1 \otimes \mathbb{H}_2)$  and  $B \in \mathbb{B}(\mathbb{H}_1)$  and  $C \in \mathbb{B}(\mathbb{H}_1)$  we have

$$\text{tr}(A(B \otimes C)) = \text{tr}_1(\text{tr}_2(A)B) \cdot \text{tr}_2(\text{tr}_1(A)C)$$

where  $\text{tr}_1$  and  $\text{tr}_2$  denote partial traces. Therefore  $\text{tr}(A(B \otimes C))$  is extreme when  $\text{tr}_1(\text{tr}_2(A)B)$  and  $\text{tr}_2(\text{tr}_1(A)C)$  are extreme. Now  $\text{tr}_1(\text{tr}_2(A)B)$  is minimal if  $B$  is supporten on the eigen space of the minimal eigenvalue of  $\text{tr}_2(A)$  and  $\text{tr}_1(\text{tr}_2(A)B)$  is maximal if  $B$  is supporten on the eigen space of the maximal eigenvalue of  $\text{tr}_2(A)$ . In particular  $B_1 \otimes C_1$  and  $B_2 \otimes C_2$  are orthogonal as elements of the set of separable states if  $B_1 \perp B_2$  or  $C_1 \perp C_2$ . Since  $\text{tr}((B_1 \otimes C_1)(B_2 \otimes C_2)) = \text{tr}(B_1 B_2) \text{tr}(C_1 C_2)$  we have

that  $B_1 \otimes C_1$  and  $B_2 \otimes C_2$  are orthogonal as elements of the set of separable states if and only if  $B_1 \otimes C_1$  and  $B_2 \otimes C_2$  are orthogonal under the Hilbert-Schmidt inner product on  $\mathbb{B}(\mathbb{H}_1 \otimes \mathbb{H}_2)$ .

Now any separable state can be decomposed into  $\sum_i \lambda_i B_i \otimes C_i$ , where  $B_i \otimes C_i$  are orthogonal pure states on  $\mathbb{B}(\mathbb{H}_1 \otimes \mathbb{H}_2)$  and since the density matrices on  $\mathbb{B}(\mathbb{H}_1 \otimes \mathbb{H}_2)$  form a spectral set the coefficients  $\lambda_i$  is uniquely determined. The separable states satisfy Bell type inequality and each Bell inequality determines a face of the set of separable states. These faces do not have complementary faces so the set of separable states does not satisfy the property called projectivity.

#### D. Symmetry

Recal that a convex set  $C$  is balanced about the origin if  $P \in C$  implies  $-P \in C$  where  $-P$  denotes the point with opposite sign of all coordinates. A spectral set of rank 2 is balanced, i.e. it is symmetric around a central point and all boundary points are extreme. Two states in the set are orthogonal if and only if they are antipodal. Any state can be decomposed into two antipodal states. If the state is not the center of the balanced set this is the only orthogonal decomposition. The center can be decomposed into a  $1/2$  and  $1/2$  mixture of any pair of antipodal points.

**Proposition 12.** *In two dimensions a simplex and a balanced set are the only types of spectral sets.*

A state space of rank 2 is said to have *symmetric transission probabilities* [7, Def. 9.2 (iii)] if for any states  $\sigma_1$  and  $\sigma_2$  there exists test  $\phi_i$  such that  $\phi_i(\sigma_i) = 1$  and  $\min_{\sigma} \phi_i(\sigma) = 0$ .

**Theorem 13.** *A spectral state space of rank 2 with symmetric transmission probabilities can be represented by a spin factor.*

*Proof:* Let  $C$  denote an intersection between the state space and a two dimensional hyperplane through the center  $c$  of the state space. Let  $\sigma_1$  and  $\sigma_2$  denote states such that  $\phi_1(\sigma_2) = \phi_2(\sigma_1) = 1/2$ . Define  $\tilde{\phi}_i = 2 \cdot \phi_i - 1$  so that  $\tilde{\phi}_i(\sigma_i) = 1$  and  $\tilde{\phi}_i(c) = 0$ . For any state  $\sigma$  we may define the coordinates by  $x = \tilde{\phi}_1(\sigma)$  and  $y = \tilde{\phi}_2(\sigma)$ . Let  $\phi$  denote a test that equals 1 on  $\sigma$  and equals 0 on the antipodal of  $\sigma$ . Let  $\tilde{\phi} = 2 \cdot \phi - 1$ . If the boundary of the state space at  $\sigma$  only has one supporting hyperplane then  $\tilde{\phi}$  is uniquely determined. Then  $\tilde{\phi}(\sigma_1) = x$  and  $\tilde{\phi}(\sigma_2) = y$ . Therefore the projection of  $\sigma_1$  along the supporting hyperplan has coordinates  $(x^2, xy)$  and the projection of  $\sigma_2$  has coordinates  $(xy, y^2)$ . Therefore the vectors  $(x^2 - 1, xy)$  and  $(xy, y^2 - 1)$  are parallel so that

$$\begin{vmatrix} x^2 - 1 & xy \\ xy & y^2 - 1 \end{vmatrix} = 0$$

$$1 - x^2 - y^2 = 0$$

so that  $\sigma$  has coordinates that lie on a unit circle. Almost all points on the boundary of  $C$  has a unique supporting hyperplane. Therefore almost all points lie on a circle so all points must lie on a circle. ■

A state space is said to be *symmetric* if any  $n$ -frame can be transformed into any other  $n$ -frame by an affine map of the state space into itself.

**Conjecture 14.** *A symmetric spectral state space is either a simplex or it can be represented as density elements of a formally real Jordan algebra.*

#### IV. QUANTIZATION OF ENERGY

Often the term quantum theory is used only for systems that have physical implementations. Here we will justify that the term “quantum” is used for Jordan algebras and even more general cases. Assume that  $S$  is finite dimensional state space. Let  $g_t : S \rightarrow S$  denote the transformation that maps a state at time 0 into state at time  $t$ . We assume that  $g_{s+t} = g_s \circ g_t$  so that we get a representation of  $\mathbb{R}_{0,+}$ . We shall restrict the discussion to the harmonic oscillator. In classical physics a harmonic oscillator has a period  $T$  such that the state at time  $t$  equals the state at time  $t + T$ . Therefore we may call a system with state space  $S$  a harmonic oscillator if  $g_{t+T} = g_t$  for all  $t \in \mathbb{R}_{0,+}$  and this representation can be extended to  $\mathbb{R}$ . Since the representation is periodic it may be considered as a representation of  $\mathbb{T} = \mathbb{R}/\tau$  where  $\tau$  is the circle constant  $2\pi$ . We say that the representation  $t \rightarrow g_t$  of  $\mathbb{T}$  into transformations  $g_t : S \rightarrow S$  is irreducible if any convex subset  $S' \subseteq S$  that is invariant under  $g_t$  spans  $S$ .

**Proposition 15.** *If  $t \rightarrow g_t$  is an irreducible representation of  $\mathbb{T}$  on the state space  $S$  then  $S$  is a point or  $S$  has the shape of a disk. If  $S$  is a disk then there exists a  $n \in \mathbb{N}$  such that  $t$  is mapped into a rotation by the angle  $n \cdot t$ .*

In the standard Hilbert space formalism the representation characterized by  $n$  is described by a Hamiltonian with energy  $\hbar\omega(n + 1/2)$ . If  $\mathbb{T}$  is represented on a finite dimensional state space but the representation is not irreducible then the representation can be decomposed into irreducible representations each characterized by a specific energy. In this sense energy is quantized. In the standard formulation of quantum mechanics the energy is represented by an operator that is an observable, but for other Jordan algebras there exists connected symmetry groups that cannot be represented by observables [8].

#### V. OPTIMIZATION

Let  $\mathcal{A}$  denote a subset of the feasible measurements such that  $a \in \mathcal{A}$  maps the convex set  $S$  into a distribution on the real numbers i.e. the distribution of a random variable. The elements of  $\mathcal{A}$  may represent feasible *actions* (decisions) that lead to a payoff like the score of a statistical decision, the energy extracted by a certain interaction with the system, (minus) the length of a codeword of the next encoded input letter using a specific code book, or the revenue of using a certain portfolio. For each  $\sigma \in S$  we define

$$\langle a, \sigma \rangle = E[a(\sigma)].$$

and

$$F(\sigma) = \sup_{a \in \mathcal{A}} \langle a, \sigma \rangle.$$

Without loss of generality we may assume that the set of actions  $\mathcal{A}$  is closed so that we may assume that there exists  $a \in \mathcal{A}$  such that  $F(\sigma) = \langle a, \sigma \rangle$  and in this case we say that  $a$  is optimal for  $\sigma$ . We note that  $F$  is convex but  $F$  need not be strictly convex.

If  $F(\sigma)$  is finite then we define the *regret* of the action  $a$  by

$$D_F(\sigma, a) = F(\sigma) - \langle a, \sigma \rangle.$$

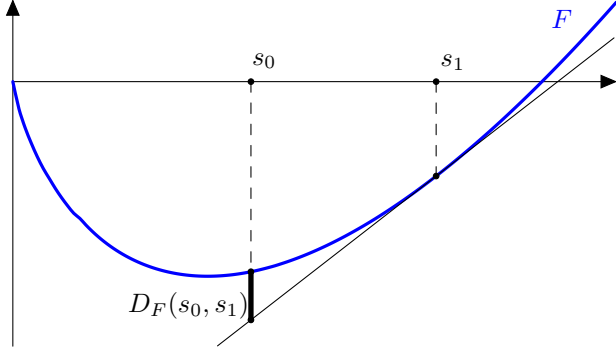


Fig. 2. The regret equals the vertical between curve and tangent.

If  $a_i$  are actions and  $(t_i)$  is a probability vector then we may define the mixed action  $\sum t_i \cdot a_i$  as the action where we do  $a_i$  with probability  $t_i$ . We note that  $\langle \sum t_i \cdot a_i, \sigma \rangle = \sum t_i \cdot \langle a_i, \sigma \rangle$ . We will assume that all such mixtures of feasible actions are also feasible. If  $a_1(\sigma) \geq a_2(\sigma)$  almost surely for all states we say that  $a_1$  dominates  $a_2$  and if  $a_1(\sigma) > a_2(\sigma)$  almost surely for all states  $\sigma$  we say that  $a_1$  strictly dominates  $a_2$ . All actions that are dominated may be removed from  $\mathcal{A}$  without changing the function  $F$ . Let  $\mathcal{A}_F$  denote the set of measurements  $m$  such that  $\langle m, \sigma \rangle \leq F(\sigma)$ . Then  $F(\sigma) = \sup_{a \in \mathcal{A}_F} \langle a, \sigma \rangle$ . Therefore we may replace  $\mathcal{A}$  by  $\mathcal{A}_F$  without changing the optimization problem.

**Definition 16.** If  $F(\sigma)$  is finite the *regret* of the action  $a$  is defined by

$$D_F(\sigma, a) = F(\sigma) - \langle a, \sigma \rangle \quad (4)$$

**Proposition 17.** The regret  $D_F$  has the following properties:

- $D_F(\sigma, a) \geq 0$  with equality if  $a$  is optimal for  $\sigma$ .
- $\sigma \rightarrow D_F(\sigma, a)$  is a convex function.
- If  $\bar{a}$  is optimal for the state  $\bar{\sigma} = \sum t_i \cdot \sigma_i$  where  $(t_1, t_2, \dots, t_\ell)$  is a probability vector then

$$\sum t_i \cdot D_F(\sigma_i, a) = \sum t_i \cdot D_F(\sigma_i, \bar{a}) + D_F(\bar{\sigma}, a).$$

- $\sum t_i \cdot D_F(\sigma_i, a)$  is minimal if  $a$  is optimal for  $\bar{\sigma} = \sum t_i \cdot \sigma_i$ .

If the state is not known exactly but we know that  $\sigma$  is one of the states  $\sigma_1, \sigma_2, \dots, \sigma_n$  then the *minimax regret* is defined as

$$C_F = \inf_a \sup_i D_F(\sigma_i, a).$$

We have the following result.

**Theorem 18.** For any set of actions

$$C_F = \sup_{\vec{t}} \inf_a \sum_i t_i \cdot D_F(\sigma_i, a)$$

where the supremum is taken over all probability vectors  $\vec{t}$  supported on  $\mathcal{S}$ .

This result can be improved.

**Theorem 19.** If  $(t_1, t_2, \dots, t_n)$  is a probability vector on the states  $\sigma_1, \sigma_2, \dots, \sigma_n$  with  $\bar{\sigma} = \sum t_i \cdot \sigma_i$  and  $a_{opt}$  is the optimal action for minimax regret then

$$C_F \geq \inf_a \sum t_i \cdot D_F(\sigma_i, a) + D_F(\bar{\sigma}, a_{opt}).$$

If  $a$  is an action and  $\sigma_{opt}$  is optimal then

$$\sup_i D_F(\sigma_i, a) \geq C_F + D_F(\sigma_{opt}, a).$$

*Proof:* See [9, Thm. 2]. ■

#### A. Bregman divergences

If the state is  $\rho$  but one acts as if the state were  $\sigma$  one suffers a regret that equals the difference between what one achieves and what could have been achieved.

**Definition 20.** If  $F(\rho)$  is finite then we define the *regret of the state*  $\sigma$  as

$$D_F(\rho, \sigma) = \inf_a D_F(\rho, a)$$

where the infimum is taken over actions  $a$  that are optimal for  $\sigma$ .

If the state  $\sigma$  has the unique optimal action  $a$  then

$$F(\rho) = D_F(\rho, \sigma) + \langle a, \rho \rangle \quad (5)$$

so the function  $F$  can be reconstructed from  $D_F$  except for an affine function of  $\rho$ . The closure of the convex hull of the set of functions  $\sigma \rightarrow \langle a, \sigma \rangle$  is uniquely determined by the convex function  $F$ .

The regret is called a *Bregman divergence* if it can be written in the following form

$$D_F(\rho, \sigma) = F(\rho) - (F(\sigma) + \langle \rho - \sigma, \nabla F(\sigma) \rangle) \quad (6)$$

where  $\langle \cdot, \cdot \rangle$  denotes some inner product. In the context of forecasting and statistical scoring rules the use of Bregman divergences dates back to [10].

**Theorem 21.** The following conditions are equivalent.

For each state  $s$  and all actions  $a_1$  and  $a_2$  such that  $F(\sigma) = \langle a_1, \sigma \rangle = \langle a_2, \sigma \rangle$  we have  $E \circ a_1 = E \circ a_2$ .

The function  $F$  is differentiable.

The regret  $D_F$  is a Bregman divergence.

We note that if  $D_F$  is a Bregman divergence and  $\sigma$  minimizes  $F$  then  $\nabla F(\sigma) = 0$  so that the formula for the Bregman divergence reduces to

$$D_F(\rho, \sigma) = F(\rho) - F(\sigma).$$

Bregman divergences satisfy the *Bregman identity*

$$\sum t_i \cdot D_F(\rho_i, \sigma) = \sum t_i \cdot D_F(\rho_i, \bar{\rho}) + D_F(\bar{\rho}, \sigma), \quad (7)$$

but if  $F$  is not differentiable this identity can be violated.

**Example 22.** Let the state space be the interval  $[0, 1]$  with two actions  $a_0(\sigma) = 1 - 2\sigma$  and  $a_1(\sigma) = 2\sigma - 1$ . Let  $\sigma_0 = 0$  and  $\sigma_1 = 1$ . Let further  $t_0 = 1/3$  and  $t_1 = 2/3$ . Then  $\bar{\sigma} = 2/3$ . If  $\sigma = 1/2$  then

$$\sum t_i \cdot D_F(\sigma_i, s) = 0$$

but

$$\begin{aligned} \sum t_i \cdot D_F(\sigma_i, \bar{\sigma}) &= \\ \frac{1}{3} \cdot (a_0(0) - a_1(0)) + \frac{2}{3} \cdot (a_1(1) - a_1(1)) &= \\ = \frac{1}{3} \cdot (1 - (-1)) = \frac{2}{3}. \end{aligned}$$

We also have  $D_F(\bar{\sigma}, \sigma) = 0$ . Clearly the Bregman identity (7) is violated.

Assume that  $\bar{a}$  is optimal for  $\bar{\rho}$ . Then

$$\begin{aligned} \sum t_i \cdot (F(\rho_i)) &= \sum t_i \cdot (D_F(\rho_i, \bar{\rho}) + \langle \bar{a}, \rho_i \rangle) \\ &= \sum t_i \cdot D_F(\rho_i, \bar{\rho}) + \left\langle \bar{a}, \sum t_i \cdot \rho_i \right\rangle \\ &= \sum t_i \cdot D_F(\rho_i, \bar{\rho}) + \langle \bar{a}, \bar{\rho} \rangle \\ &= \sum t_i \cdot D_F(\rho_i, \bar{\rho}) + F(\bar{\rho}). \end{aligned}$$

Therefore

$$\sum t_i \cdot D_F(\rho_i, \bar{\rho}) = \sum t_i \cdot (F(\rho_i)) - F(\bar{\rho}).$$

## VI. SUFFICIENCY CONDITIONS

In this section we will introduce various conditions on a Bregman divergence. Under some mild conditions they turn out to be equivalent.

**Theorem 23.** Let  $D_F$  denote a regret function defined on an arbitrary state space. Then we have the implications 3.  $\Rightarrow$  4.  $\Rightarrow$  5.  $\Rightarrow$  6.  $\Rightarrow$  1.  $\Rightarrow$  2.

1. The function  $F$  equals entropy times a constant plus an affine function.

2. The regret  $D_F$  is proportional to information divergence.

3. The regret function satisfies strong sufficiency.

4. The regret is monotone, i.e. it satisfies the data processing inequality.

5. The regret is sufficiency stable.

6. The regret is local.

The conditions 1., 2., and 6. are equivalent on formally real Jordan algebras with at least three orthogonal states..

A. Entropy in Jordan algebras

**Definition 24.** Let  $x$  denote an element in a positive cone. The entropy of  $x$  is defined as

$$H(x) = \inf \left( - \sum_{i=1}^n \lambda_i \ln(\lambda_i) \right)$$

where the infimum is taken over all spectra of  $x$ .

Since entropy is decreasing under majorization the entropy of  $x$  is attained at an orthogonal decomposition. This definition extends a similar definition of the entropy of a state as defined by Uhlmann [11].

In general this definition of entropy does not provide a concave function on the positive cone. For instance the entropy of points in the square from Example 6 has local maximum in the four points with maximal spectrum in the majorization ordering. A characterization of the convex sets with concave entropy functions is lacking.

**Proposition 25.** Assume that the entropy function  $H$  on a state space is strictly concave. Then any uniform mixture of orthogonal pure states give the maximum entropy state.

*Proof:* Assume that  $\sigma = \frac{1}{n} \sum_{i=1}^n \sigma_i$  and  $\rho = \frac{1}{n} \sum_{i=1}^n \rho_i$  are decompositions into orthogonal pure states. Then  $H(\sigma) = H(\rho) = \ln(n)$  and  $H\left(\frac{\sigma+\rho}{2}\right) \geq \ln(n)$  with equality, which implies that  $\sigma = \rho$  ■

The entropy is defined as for general convex set and we will prove that  $H$  is a concave on the cone of positive elements in a Jordan algebra. The following exposition is inspired by similar result for complex matrix algebras stated in [12], but the proofs have been changed so that they are valid on Jordan algebras.

**Lemma 26.** For elements  $A$  and  $B$  in a Jordan algebra and any analytic function  $f$  we have

$$\left. \frac{d}{dt} \text{tr}[f(A + tB)] \right|_{t=0} = \text{tr}[f'(A) \circ B].$$

*Proof:* First assume that  $f(z) = z^r$ . Then

$$\begin{aligned} \left. \frac{d}{dt} \text{tr}[f(A + tB)] \right|_{t=0} &= \text{tr} \left[ \left. \frac{d}{dt} (A + tB)^r \right|_{t=0} \right] \\ &= \text{tr} \left[ \sum_{i=0}^{r-1} A^i \circ B \circ A^{r-1-i} \right] \\ &= \text{tr} \left[ \sum_{i=0}^{r-1} A^i \circ A^{r-1-i} \circ B \right] \\ &= \text{tr} [n \cdot A^{r-1} \circ B] \\ &= \text{tr}[f'(A) \circ B]. \end{aligned}$$

As a consequence the theorem holds for any polynomial and also for any analytic function because such functions can be approximated by polynomials. ■

**Lemma 27.** In a real Jordan algebra the following formula holds for any analytic function  $f$

$$\left. \frac{d^2}{dt^2} \text{tr}[f(A + tB)] \right|_{t=0} = \sum_{k, \ell} a_{k, \ell} \text{tr}[(E_k \circ B) \circ (E_\ell \circ B)]$$

where  $A = \sum_k \lambda_k E_k$  is an orthogonal decomposition and

$$a_{k, \ell} = \begin{cases} \frac{f'(\lambda_k) - f'(\lambda_\ell)}{\lambda_k - \lambda_\ell} & \text{for } \lambda_k \neq \lambda_\ell, \\ f''(\lambda_k) & \text{for } \lambda_k = \lambda_\ell. \end{cases}$$

*Proof:* Let  $f$  denote the function  $z^r$ . Then

$$\begin{aligned} \frac{d^2}{dt^2} \text{tr} [f(A + tB)] \Big|_{t=0} &= \frac{d}{dt} \text{tr} [f'(A + tB) \circ B] \Big|_{t=0} \\ &= \frac{d}{dt} \text{tr} \left[ r \cdot (A + tB)^{r-1} \circ B \right] \Big|_{t=0} \\ &= r \cdot \text{tr} \left[ \sum_{j=0}^{r-2} \left( ((A^j \circ B) \circ \overbrace{A \circ A \circ \dots \circ A}^{r-j-2 \text{ times}}) \circ B \right) \right] \\ &= r \cdot \text{tr} \left[ \sum_{j=0}^{r-2} (A^j \circ B) \circ (A^{r-2-j} \circ B) \right]. \end{aligned}$$

Assume  $A = \sum_k \lambda_k E_k$  where  $E_k$  are orthogonal pure states. Then

$$\begin{aligned} \frac{d^2}{dt^2} \text{tr} [f(A + tB)] \Big|_{t=0} &= \\ r \cdot \text{tr} \left[ \sum_{j=0}^{r-2} \left( \sum_k \lambda_k^j \cdot E_k \circ B \right) \circ \left( \sum_\ell \lambda_\ell^{r-2-j} \cdot E_\ell \circ B \right) \right] \\ &= r \cdot \sum_{j=0}^{r-2} \sum_{k,\ell} \lambda_k^j \lambda_\ell^{r-2-j} \text{tr} [(E_k \circ B) \circ (E_\ell \circ B)] \\ &= r \cdot \sum_{k,\ell} \left( \sum_{j=0}^{r-2} \lambda_k^j \lambda_\ell^{r-2-j} \right) \text{tr} [(E_k \circ B) \circ (E_\ell \circ B)] \\ &= \sum_{k,\ell} \left( r \cdot \sum_{j=0}^{r-2} \lambda_k^j \lambda_\ell^{r-2-j} \right) \text{tr} [(E_k \circ B) \circ (E_\ell \circ B)]. \end{aligned}$$

Now

$$\begin{aligned} \sum_{j=0}^{r-2} \lambda_k^j \lambda_\ell^{r-2-j} &= \begin{cases} \frac{r \cdot \lambda_k^{r-1} - r \cdot \lambda_\ell^{r-1}}{\lambda_k - \lambda_\ell} & \text{for } \lambda_k \neq \lambda_\ell, \\ r \cdot (r-1) \lambda_k^{r-2} & \text{for } \lambda_k = \lambda_\ell \end{cases} \\ &= \begin{cases} \frac{f'(\lambda_k) - f'(\lambda_\ell)}{\lambda_k - \lambda_\ell} & \text{for } \lambda_k \neq \lambda_\ell, \\ f''(\lambda_k) & \text{for } \lambda_k = \lambda_\ell. \end{cases} \end{aligned}$$

Since the formula holds of all powers, it also holds for all polynomials and for all analytic functions because these can be approximated by polynomials. ■

**Theorem 28.** *In a formally real Jordan algebra the entropy function is a concave function on the positive cone.*

*Proof:* Let  $f$  denote the holomorphic function  $f(z) = -z \ln z$ ,  $z > 0$ . We have to prove that  $\text{tr} [f((1-t)A + tX)] = \text{tr} [f(A + tB)]$  is concave where  $B = X - A$ . We have

$$\frac{d^2}{dt^2} \text{tr} [f(A + tB)] \Big|_{t=0} = \sum_{k,\ell} a_{k,\ell} \text{tr} [(E_k \circ B) \circ (E_\ell \circ B)]$$

and the coefficients  $a_{k,\ell}$  are negative because  $f$  is concave. We need to prove that  $\text{tr} [(E_k \circ B) \circ (E_\ell \circ B)] \geq 0$ , but a formally real Jordan algebra can be written as a sum of simple Jordan algebras so it is sufficient to prove positivity on simple algebras. ■

Assume that the Jordan algebra is special we have

$$\begin{aligned} \text{tr} [(E_k \circ B) \circ (E_\ell \circ B)] &= \frac{1}{4} \text{tr} [(E_k B + B E_k) (E_\ell B + B E_\ell)] \\ &= \frac{1}{4} \text{tr} [E_k B E_\ell B + E_k B^2 E_\ell + B E_k E_\ell B + B E_k B E_\ell] \\ &= \frac{1}{4} \text{tr} [(E_k B E_\ell) (E_k B E_\ell)^* + (E_\ell B E_k) (E_\ell B E_k)^*] \\ &\geq 0. \end{aligned}$$

Assume that the Jordan algebra is exceptional, i.e. the Albert algebra. There exists an  $F_4$  automorphism of the Jordan algebra such that  $E_k$  is orthogonal and without loss of generality we may assume that

$$E_k = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The idempotent  $E_\ell$  has the form  $vv^*$  for some column vector  $v$  with entries in a quaternionic sub-algebra. Assume that  $v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$ . Then

$$E_\ell = \begin{pmatrix} v_1 \bar{v}_1 & v_1 \bar{v}_2 & v_1 \bar{v}_3 \\ v_2 \bar{v}_1 & v_2 \bar{v}_2 & v_2 \bar{v}_3 \\ v_3 \bar{v}_1 & v_3 \bar{v}_2 & v_3 \bar{v}_3 \end{pmatrix}.$$

Now  $\text{tr} (E_k \circ E_\ell) = v_1 \bar{v}_1 = 0$  since  $E_k$  and  $E_\ell$  are orthogonal. Therefore  $v_1 = 0$ . Without loss of generality we may assume that

$$\begin{aligned} E_\ell &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\ B &= \begin{pmatrix} p & a & \bar{b} \\ \bar{a} & m & c \\ b & \bar{c} & n \end{pmatrix}. \end{aligned}$$

Then

$$\begin{aligned} E_k \circ B &= \frac{1}{2} \begin{pmatrix} 2p & a & \bar{b} \\ \bar{a} & 0 & 0 \\ b & 0 & 0 \end{pmatrix} \\ E_\ell \circ B &= \frac{1}{2} \begin{pmatrix} 0 & a & 0 \\ \bar{a} & 2m & c \\ 0 & \bar{c} & 0 \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} (E_k \circ B) \circ (E_\ell \circ B) &= \\ \frac{1}{4} \begin{pmatrix} 2a\bar{a} & 2pa + 2ma + \bar{b}\bar{c} & ac \\ 2p\bar{a} + 2m\bar{a} + \bar{c}\bar{b} & 2\bar{a}a & \bar{a}\bar{b} \\ \bar{c}\bar{b} & ba & 0 \end{pmatrix}. \end{aligned}$$

Therefore

$$\text{tr} [(E_k \circ B) \circ (E_\ell \circ B)] = |a|^2 \geq 0.$$

## B. Information divergence

**Definition 29.** If the entropy is a concave function then the Bregman divergence  $D_{-H}$  is called *information divergence*.

The information divergence is also called *Kullback-Leibler divergence*, *relative entropy* or *quantum relative entropy*. In a Jordan algebra we get

$$\begin{aligned} D_{-H}(P, Q) &= \\ &= -H(P) - (-H(Q) + \langle P - Q, -\nabla H(Q) \rangle) \\ &= H(Q) - H(P) + \langle P - Q, \nabla H(Q) \rangle \\ &= \text{tr}[f(Q) - \text{tr}(f(P)) + \text{tr}((P - Q) \circ f'(Q))] \\ &= \text{tr}[f(Q) - f(P) + (P - Q) \circ f'(Q)] \end{aligned}$$

where  $f(x) = -x \ln(x)$ . Now  $f'(x) = -\ln(x) - 1$  so that

$$\begin{aligned} f(Q) - f(P) + (P - Q) f'(Q) &= \\ &= -Q \circ \ln(Q) + P \circ \ln(P) + (P - Q)(-\ln(Q) - 1) \\ &= P \circ (\ln(P) - \ln(Q)) + Q - P. \end{aligned}$$

Hence

$$D_{-H}(P, Q) = \text{tr}[P \circ (\ln(P) - \ln(Q)) + Q - P]$$

and for states  $P, Q$  it reduces to

$$D_{-H}(P, Q) = \text{tr}[P \circ \ln(P) - P \circ \ln(Q)].$$

**Proposition 30.** If a state space  $\mathcal{S}$  has rank 2 and the entropy function  $H$  is concave then the state space is spectral.

*Proof:* Let  $C_{-H}$  denote the capacity of the state space and. We have

$$\begin{aligned} C_{-H} &= \sup_{(q_1, q_2, \dots, q_n)} \sum_j q_j D \left( \rho_j \left\| \sum_i q_i \rho_i \right\| \right) \\ &= \sup_{(q_1, q_2, \dots, q_n)} H \left( \sum_i q_i \rho_i \right) - \sum_j q_j H(\rho_j) \\ &= \sup_{\rho \in \mathcal{S}} H(\rho). \end{aligned}$$

If  $\rho = q_1 \rho_1 + q_2 \rho_2$  then  $H(\rho) \leq -q_1 \ln(q_1) - q_2 \ln(q_2) \leq \ln(2)$ . Therefore  $\sup_{\rho \in \mathcal{S}} H(\rho) \leq \ln(2)$ . Let  $\bar{\sigma}$  denote a capacity achieving state. Assume that  $\bar{\sigma} = p_1 \sigma_1 + p_2 \sigma_2$  where  $\sigma_1$  and  $\sigma_2$  are pure orthogonal states. Then

$$\begin{aligned} C_{-H} &= \sup_{\rho \in \mathcal{S}} D(\rho \| \bar{\sigma}) \\ &\geq \max_{i=1,2} \{D(\sigma_i \| \bar{\sigma}), D(\sigma_i \| \bar{\sigma})\} \\ &= \max_{i=1,2} \left\{ \ln \left( \frac{1}{p_1} \right), \ln \left( \frac{1}{p_2} \right) \right\} \\ &\geq \ln(2). \end{aligned}$$

The minimax-result then implies that  $C_{-H} = \ln(2)$ . Therefore  $\bar{\sigma} = \frac{1}{2} \sigma_1 + \frac{1}{2} \sigma_2$ . If  $\rho_1$  and  $\rho_2$  are orthogonal and  $\bar{\rho} = \frac{1}{2} \rho_1 + \frac{1}{2} \rho_2$ . Then

$$\begin{aligned} \ln(2) &\geq \frac{1}{2} D(\rho_1 \| \rho) + \frac{1}{2} D(\rho_2 \| \bar{\rho}) + D(\bar{\rho} \| \bar{\sigma}) \\ &= \frac{1}{2} \ln(2) + \frac{1}{2} \ln(2) + D(\bar{\rho} \| \bar{\sigma}) \\ &= \ln(2) + D(\bar{\rho} \| \bar{\sigma}). \end{aligned}$$

Therefore  $D(\bar{\rho} \| \bar{\sigma}) = 0$  so that  $\bar{\rho} = \bar{\sigma}$ . Therefore the state space is balanced and thereby it is spectral. ■

## C. Strong monotonicity

**Definition 31.** A regret function  $D_F$  is said to satisfy *strong monotonicity* if for any transformation  $\Phi$  of the state space into itself the equation

$$D_F(\Phi(\rho), \Phi(\sigma)) = D_F(\rho, \sigma)$$

implies that there exists a recovery map  $\Psi$ , i.e. a map of the state space into itself such that  $\Psi(\Phi(\rho)) = \rho$  and  $\Psi(\Phi(\sigma)) = \sigma$ .

In statistics where the state space is a simplex strong monotonicity is well established.

**Example 32.** Squared Euclidean distance on a spin factor is a Bregman divergence that satisfies strong monotonicity. To see this we note an transformation  $\Phi$  can be decomposed into a translation and a linear map. Since the transformation maps the unit ball into itself the maximal eigenvalue of the linear map must be 1. Therefore

$$D_F(\Phi(\rho), \Phi(\sigma)) = D_F(\rho, \sigma)$$

implies that  $\rho$  and  $\sigma$  belong to a subspace that has eigenvalue 1. The intersection of the subspace spanned by  $\rho$  and  $\sigma$  and the state space is a disc of radius 1 and this disc must be mapped into another disc of radius 1 in the state space. Since any disc of radius 1 can be mapped into any other disc of radius one there exists a recovery map.

**Proposition 33.** If the regret function  $D_F$  is strongly monotone then  $F$  is strictly convex.

*Proof:* Assume that  $F$  is not strictly convex. Then there exists states  $\rho$  and  $\sigma$  such that  $F\left(\frac{\rho+\sigma}{2}\right) = \frac{F(\rho)+F(\sigma)}{2}$ . Then  $D_F(\rho, \sigma) = 0$ . Let  $\Phi$  denote a contraction around  $\frac{\rho+\sigma}{2}$ . Then  $D_F(\Phi(\rho), \Phi(\sigma)) = 0$  but a contraction cannot have a recovery map on a compact set. ■

For density matrices over the complex numbers strong monotonicity has been proved for completely positive maps in [13]. Some new results on this topic can be found in [13].

## D. Feasible transformations and monotonicity

We consider a set  $\mathcal{T}$  of *feasible transformations* of the state space. By a feasible transformation we mean a transformation that we are able to perform on the state space before we choose a feasible action. Let  $\Phi : \mathcal{S} \rightarrow \mathcal{S}$  denote a feasible transformation and let  $a$  denote a feasible action. Then  $a \circ \Phi$  is the action



$\sigma \rightarrow a(\Phi(\sigma))$ . Thus the set of feasible transformations acts on the set of actions. If  $\Psi$  and  $\Phi$  are feasible transformations then we will assume that  $\Psi \circ \Phi$  is also feasible. Further we will assume that the identity is feasible. Let  $\mathcal{F}$  denote the monoid of feasible transformations. Finally we will assume that  $(1-s) \cdot \Psi + s \cdot \Phi$  is feasible for  $s \in [0, 1]$  so that  $\mathcal{F}$  becomes a convex monoid.

**Proposition 34** (The principle of lost opportunities). *If  $\Phi$  is a feasible transformation then*

$$F(\Phi(\sigma)) \leq F(\rho). \quad (8)$$

*Proof:* See [9, Prop. 4]. ■

Since the feasible transformations increase the value of  $F$  the set of states with minimal value of  $F$  is invariant under feasible transformations.

**Proposition 35.** *Let  $S$  be a state space. Then the set  $\mathcal{T}_F$  of transformations  $\Phi : S \curvearrowright S$  such that  $F(\Phi(\sigma)) \leq F(\sigma)$  for all  $\sigma \in S$  is a convex monoid and for any action  $a \in \mathcal{A}_F$  we have that  $a \circ \Phi \in \mathcal{A}_F$ .*

*Proof:* See [9, Prop. 5]. ■

**Corollary 36** (Semi-monotonicity). *Let  $\Phi$  denote a feasible transformation and let  $\sigma$  denote a state that minimizes the function  $F$ . If  $D_F$  is a Bregman divergence then*

$$D_F(\Phi(\rho), \Phi(\sigma)) \leq D_F(\rho, \sigma). \quad (9)$$

*Proof:* See [9, Cor. 1]. ■

Next we introduce the stronger notion of monotonicity.

**Definition 37.** Let  $D_F$  denote a regret function on the convex set  $C$ . Then  $D_F$  is said to be *monotone* if

$$D_F(\Phi(\rho), \Phi(\sigma)) \leq D_F(\rho, \sigma)$$

for any affine transformation  $\Phi : C \rightarrow C$ .

In information theory an inequality of this type is often called a *data processing inequality*. In general a regret function need not be monotone [9, Ex. 5]. Recently it has been proved that information divergence on a complex Hilbert space is decreasing under positive trace preserving maps [14], [15]. Previously this was only known to hold if some extra condition like complete positivity was assumed.

**Theorem 38.** *Information divergence is monotone under any positive trace preserving map on a special Jordan algebra.*

*Proof:* The proof is a step by step repetition of the proof by Müller-Hermes and Reep [14], where they proved the theorem for density matrices over the complex numbers. See also [15] where the same proof technique is used. In their proof they use the *sandwiched Rényi divergence* defined by  $D_\alpha(\rho \parallel \sigma) = \frac{1}{\alpha-1} \ln \left( \text{tr} \left[ \sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}} \right]^\alpha \right)$ , and we note that this quantity can be defined and manipulated as in the proof of Reep and Müller-Hermes as long as the algebra is associative. ■

**Proposition 39.** *If a regret function is strongly monotone, then the regret function is monotone.*

*Proof:* Assume that the regret function  $D_F$  is strongly monotone. Then there exists states  $\rho$  and  $\sigma$  and a transformation  $\Phi$  such that  $D_F(\Phi(\rho), \Phi(\sigma)) \geq D_F(\rho, \sigma)$ . Let  $\Theta_\epsilon$  denote a contraction around  $\Phi(\rho)$  with factor  $\epsilon \in [0, 1]$ . Then there exists a value of  $\epsilon$  such that  $D_F(\Theta_\epsilon(\Phi(\rho)), \Theta_\epsilon(\Phi(\sigma))) = D_F(\rho, \sigma)$ . Therefore  $\Theta_\epsilon \circ \Phi$  has a reverse  $\Psi$  such that

$$\Psi((\Theta_\epsilon \circ \Phi)(\rho)) = \rho$$

$$\Psi((\Theta_\epsilon \circ \Phi)(\sigma)) = \sigma$$

Therefore

$$(\Phi \circ \Psi)(\Theta_\epsilon(\Phi(\rho))) = \Phi(\rho)$$

$$(\Phi \circ \Psi)(\Theta_\epsilon(\Phi(\sigma))) = \Phi(\sigma)$$

so that  $\Phi \circ \Psi$  is a recovery map of the contraction  $\Theta_\epsilon$ . This is only possible if  $\epsilon = 1$  implying that  $D_F(\Phi(\rho), \Phi(\sigma)) = D_F(\rho, \sigma)$ . ■

### E. Sufficiency

The present definition of sufficiency is based on [16], but there are a number of other equivalent ways of defining this concept. We refer to [13] where the notion of sufficiency is discussed in great detail.

**Definition 40.** Let  $(\sigma_\theta)_\theta$  denote a family of states and let  $\Phi$  denote an affine transformation  $S \rightarrow T$  where  $S$  and  $T$  denote state spaces. Then  $\Phi$  is said to be *sufficient* for  $(\sigma_\theta)_\theta$  if there exists an affine transformation  $\Psi : T \rightarrow S$  such that  $\Psi(\Phi(\sigma_\theta)) = \sigma_\theta$ . We say that  $\Phi$  is *reversible* if  $\Phi$  is feasible and there exist a feasible  $\Psi$  such that  $\Psi(\Phi(\sigma_\theta)) = \sigma_\theta$ .

**Proposition 41.** *If  $D_F$  is a regret function and  $\Phi$  is reversible for  $\rho$  and  $\sigma$  then*

$$D_F(\Phi(\rho), \Phi(\sigma)) = D_F(\rho, \sigma).$$

*Proof:* See [9, Prop. 6]. ■

The notion of sufficiency as a property of divergences was introduced in [17]. The crucial idea of restricting the attention to transformations of the state space into itself was introduced in [18]. It was shown in [18] that a Bregman divergence on the simplex of distributions on an alphabet that is not binary determines the divergence except for a multiplicative factor. Here we generalize the notion of sufficiency from Bregman divergences to regret functions.

**Definition 42.** We say that the regret  $D_F$  on the state space  $S$  satisfies *sufficiency* if  $D_F(\Phi(\rho), \Phi(\sigma)) = D_F(\rho, \sigma)$  for any affine transformation  $S \rightarrow S$  that is sufficient for  $(\rho, \sigma)$ .

**Proposition 43.** *A monotone regret function  $D_F$  satisfies sufficiency.*

*Proof:* See [9, Prop. 7]. ■

Combining the previous results we get that information divergence on a special Jordan algebra satisfies sufficiency.

**Lemma 44.** *If a regret function on a state space of rank 2 satisfies sufficiency, then the state space is spectral.*

*Proof:* For  $i = 1, 2$  let  $\sigma_i$  and  $\rho_i$  denote pure states and let  $\phi_i$  is a test such that  $\phi_i(\sigma_i) = 1$  and  $\phi_i(\rho_i) = 0$ . Let  $m_i$  denote the midpoint  $m_i = \frac{1}{2}\sigma_i + \frac{1}{2}\rho_i$ . Then the transformation  $\Phi$  given by

$$\Phi(\pi) = \phi_1(\pi)\sigma_2 + (1 - \phi_1(\pi))\rho_2$$

is sufficient for the pair  $(\sigma_1, m_1)$  with recovery map  $\Psi$  given by

$$\Psi(\pi) = \phi_2(\pi)\sigma_1 + (1 - \phi_2(\pi))\rho_1.$$

Using sufficiency we have  $D_F(p\sigma_1 + (1-p)\rho_1, m_1) = D_F(p\sigma_2 + (1-p)\rho_2, m_2)$ . Define  $f(p) = D_F(p\sigma_1 + (1-p)\rho_1, m_1)$ . Then  $D_F(p\sigma_2 + (1-p)\rho_2, m_2) = f(p)$  for any pair of orthogonal states  $(\sigma_2, \rho_2)$  so this divergence is completely determined by the spectrum  $(p, 1-p)$ . In particular any orthogonal decomposition must have the same spectrum so that the state space is spectral. ■

Let  $\mathcal{S}$  denote a state space of rank two with regret function that satisfies sufficiency. Then  $\mathcal{S}$  is spectral and therefore balanced. Let  $c$  denote the center of  $\mathcal{S}$ . Then any state  $\sigma$  has a decomposition  $\sigma = t\rho + (1-t)c$  where  $\rho$  is a pure state. We have  $D_F(\sigma, c) = f(\frac{t+1}{2})$ . Then there exists an affine function  $g$  such that  $F(\sigma) = f(\frac{t+1}{2}) + g(t)$ . For  $t = 0$  we get  $g(0) = F(c)$ . Therefore  $F(\sigma) = f(\frac{t+1}{2}) + F(c) + h(t)$  where  $h$  is a linear function that that may depend on  $\rho$ .

**Proposition 45.** *A monotone regret function is a Bregman divergence.*

*Proof:* A regret function is given by a function  $F$  and we have to prove that this function is differentiable. Since  $F$  is convex it is sufficient to prove that  $F$  is differentiable in any direction. Let  $\rho$  and  $\sigma$  denote two states and let  $\Phi$  denote a contraction around  $\rho$  by a factor  $t \in [0, 1]$ . Then  $\Phi(\rho) = \rho$  and  $\Phi(\sigma) = (1-t)\rho + t\sigma$ . Let  $\pi$  denote a state on the line between  $\rho$  and  $\sigma$ . Monotonicity implies that

$$D_F(\Phi(\sigma), \Phi(\pi)) \leq D_F(\sigma, \pi)$$

$$D_F((1-t)\rho + t\sigma, (1-t)\rho + t\pi) \leq D_F(\sigma, \pi)$$

$$\lim_{t \rightarrow 1-} D_F((1-t)\rho + t\sigma, (1-t)\rho + t\pi) \leq D_F(\sigma, \pi)$$

$$\begin{aligned} \lim_{t \rightarrow 1-} \left( \inf_a F((1-t)\rho + t\sigma) - \langle a, (1-t)\rho + t\pi \rangle \right) \\ \leq \inf_a F(\sigma) - \langle a, \sigma \rangle \end{aligned}$$

$$F(\sigma) - \lim_{t \rightarrow 1-} \left( \sup_a \langle a, (1-t)\rho + t\sigma \rangle \right) \leq F(\sigma) - \sup_a \langle a, \sigma \rangle$$

where  $a$  on the left side should be optimal for  $\Phi(\pi)$  and  $a$  on the right hand side should be optimal for  $\pi$ . Then

$$\lim_{t \rightarrow 1-} \left( \sup_a \langle a, (1-t)\rho + t\sigma \rangle \right) \geq \sup_a \langle a, \sigma \rangle$$

which implies that the left derivative and the right derivative of  $F$  at  $\pi$  are the same so that  $F$  is differentiable at  $\pi$ .

Another way to state the result is that a regret function that is based on a function  $F$  that is not differentiable violates monotonicity. ■

**Theorem 46.** *If a state space of rank 2 has a monotone regret function, then the state space can be represented by a spin factor.*

*Proof:* First we note that a monotone regret function is a Bregman divergence. Since a monotone Bregman divergence satisfies sufficiency the state space must be balanced. Let  $c$  denote the center of the state space. Without loss of generality we will assume that the state space  $\mathcal{S}$  is two dimensional. Let  $\Phi$  denote a transformation of the state space that first rotate it by an angle of  $\theta$  and then dilate it around  $c$  with a factor  $r < 1$ . The factor can be chosen in such a way that the boundary of  $\Phi(\mathcal{S})$  touches the boundary of  $\mathcal{S}$ . Assume that the boundary of  $\Phi(\mathcal{S})$  touches the boundary of  $\mathcal{S}$  in the pure state  $\sigma'$  and let  $\phi$  denote a test on  $\mathcal{S}$  such that  $\phi(\sigma') = 1$  and  $\phi(c) = 1/2$ . Let  $\Psi$  denote a map that stretches  $\Phi(\mathcal{S})$  without changing  $\phi$ . The map  $\Psi$  should stretch so much that the boundary of  $\Psi(\Phi(\mathcal{S}))$  touch the boundary of  $\mathcal{S}$  in a pure state  $\rho'$  different  $\rho'$  that is not co-linear with  $c$  and  $\sigma'$ . Let  $\rho$  and  $\sigma$  denote the preimages of  $\rho'$  and  $\sigma'$ . Let  $\pi$  denote a pure of the shortest path along the boundary of  $\mathcal{S}$ . Then there exist a mixture  $\bar{\sigma} = (1-s)\pi + s\sigma$  that also has a decomposition of the form  $\bar{\sigma} = (1-t)\rho + t\sigma$ .

Then

$$\bar{\sigma}' = (1-s)\pi' + s\sigma' = (1-t)\rho' + t\sigma'.$$

Using monotonicity we have

$$\begin{aligned} (1-t) \cdot D_F(\rho', \bar{\sigma}') + t \cdot D_F(\sigma', \bar{\sigma}') \\ \leq (1-t) \cdot D_F(\rho, \bar{\sigma}) + t \cdot D_F(\sigma, \bar{\sigma}) \end{aligned}$$

so that

$$\begin{aligned} (1-t) \cdot D_F(\rho', c) + t \cdot D_F(\sigma', c) - D_F(\bar{\sigma}', c) \\ \leq (1-t) \cdot D_F(\rho, c) + t \cdot D_F(\sigma, c) - D_F(\bar{\sigma}, c) \end{aligned}$$

and  $D_F(\bar{\sigma}', c) \geq D_F(\bar{\sigma}, c)$ . Since  $\bar{\sigma} = (1-s)\pi + s\sigma$  and  $\bar{\sigma}' = (1-s)\pi' + s\sigma'$  and  $\pi$  is a pure state  $\pi'$  must also be a pure state. Therefore  $\Psi \circ \Phi$  maps pure states into pure states so that  $\Psi \circ \Phi$  must be a bijection. Since any map of the state space into itself can be stretched into a bijection the state space must be a disc. ■

#### F. Locality

Often it is relevant to use the following weak version of the sufficiency property.

**Definition 47.** The regret function  $D_F$  is said to be local if

$$D_F(\rho, (1-t)\rho + t\sigma_1) = D_F(\rho, (1-t)\rho + t\sigma_2)$$

when  $\rho, \sigma_i$  are perfectly distinguishable for  $i = 1, 2$  and  $t \in ]0, 1[$ .

**Example 48.** On a state space of rank 2 any regret function  $D_F$  is local. The reason is that if  $\sigma_1$  and  $\sigma_2$  are states that are orthogonal to  $\rho$  then  $\sigma_1 = \sigma_2$ .

**Proposition 49.** A regret function  $D_F$  that satisfies sufficiency on a convex set is local.

*Proof:* [9, Prop. 8]. ■

**Proposition 50.** Let  $C$  denote a spectral convex set with a concave entropy function  $H$ . Then information divergence  $D_{-H}$  satisfies locality.

*Proof:* Assume that  $\rho$  and  $\sigma$  are two perfectly distinguishable states. Then one can make decompositions

$$\begin{aligned}\rho &= \sum_i p_i \cdot \rho_i \\ \sigma &= \sum_j q_j \cdot \sigma_j.\end{aligned}$$

Then

$$\begin{aligned}D_{-H}(\rho, (1-t)\rho + t\sigma) &= \sum_i p_i \cdot \ln \frac{p_i}{(1-t)p_i} \\ &= \sum_i p_i \cdot \ln \frac{1}{1-p} \\ &= \ln \frac{1}{1-p},\end{aligned}$$

which does not depend on  $\sigma$  as long as  $\rho, \sigma$  are perfectly distinguishable. ■

**Proposition 51.** Let  $C$  denote a convex set of rank 2. Then any regret function is local.

The following lemma follows from Alexandrov's theorem. See [19, Theorem 25.5] for details.

**Lemma 52.** A convex function on a finite dimensional convex set is differentiable almost everywhere with respect to the Lebesgue measure.

**Theorem 53.** Let  $S$  be a convex set of rank  $r \geq 3$  and assume that  $S$  is weakly spectral. If a regret function  $D_F$  defined on  $S$  is local then the state space  $S$  is spectral and the regret is a Bregman divergence generated by the entropy times some constant.

*Proof:* In the following proof we will assume that the regret function is based on the convex function  $F : S \rightarrow \mathbb{R}$ .

Let  $K$  denote the convex hull of a set  $\sigma_1, \sigma_2, \dots, \sigma_r$  of orthogonal states. Let  $f_i$  denote the function  $f_i(x) = D_F(\sigma_i, x\sigma_i + (1-x)\sigma_j)$ . Since  $D_F$  is local we have  $f_i(x) = D_F(\sigma_i, x\sigma_i + (1-x)\sigma_{i+1})$  for any  $j \neq i$ . Note that  $f_i$  is decreasing and continuous from the left. Let  $\rho = \sum p_i \sigma_i$  and  $\sigma = \sum q_i \sigma_i$ . If  $F$  is differentiable in  $\rho$  then locality implies that

$$\begin{aligned}D_F(\rho, \sigma) &= \sum p_i D_F(\sigma_i, \sigma) - \sum p_i D_F(\sigma_i, \rho) \\ &= \sum p_i f_i(q_i) - \sum p_i f_i(p_i) \\ &= \sum p_i (f_i(q_i) - f_i(p_i)).\end{aligned}$$

Note that  $\rho \rightarrow D_F(\rho, \sigma)$  is a convex function and thereby it is continuous. Assume that  $\rho_0$  is an arbitrary element in  $K$

and let  $(\rho_n)_{n \in \mathbb{N}}$  denote a sequence such that  $\rho_n \rightarrow \rho_0$  for  $n \rightarrow \infty$ . The sequence  $(\rho_n)_{n \in \mathbb{N}}$  can be chosen so that regret is differentiable in  $\rho_n$  for all  $n \in \mathbb{N}$ . Further the sequence  $\rho_n$  can be chosen such that  $p_{n,i}$  is increasing for all  $i \neq j$ . Then

$$\begin{aligned}D_F(\rho_0, \sigma) &= \sum p_i (f_i(q_i) - f_i(p_i)) \\ &\quad + p_{0,j} f_j(p_{0,j}) - p_{0,j} \lim_{n \rightarrow \infty} f_j(p_{n,j}).\end{aligned}$$

Similarly, if the sequence  $\rho_n$  can be chosen such that  $p_{n,i}$  is increasing for all  $i \neq j, j+1$  then

$$\begin{aligned}D_F(\rho_0, \sigma) &= \sum p_i (f_i(q_i) - f_i(p_i)) + p_{0,j} f_j(p_{0,j}) - p_{0,j} \lim_{n \rightarrow \infty} f_j(p_{n,j}) \\ &\quad + p_{0,j+1} f_{j+1}(p_{0,j+1}) - p_{0,j+1} \lim_{n \rightarrow \infty} f_{j+1}(p_{n,j+1}),\end{aligned}$$

which implies that  $p_{0,j+1} f_{j+1}(p_{0,j+1}) - p_{0,j+1} \lim_{n \rightarrow \infty} f_{j+1}(p_{n,j+1}) = 0$  and that  $\lim_{n \rightarrow \infty} f_{j+1}(p_{n,j+1}) = f_{j+1}(p_{0,j+1})$  for all  $j$  so that

$$D_F(\rho_0, \sigma) = \sum p_i (f_i(q_i) - f_i(p_i))$$

even if the regret is not differentiable in  $\rho_0$ .

As a function of  $\sigma$  the regret has minimum when  $\sigma = \rho$ . We have

$$x(f_i(y) - f_i(x)) + z(f_j(w) - f_j(z)) \geq 0$$

where  $x + z = y + w$ . We also have

$$x(f_j(y) - f_j(x)) + z(f_i(w) - f_i(z)) \geq 0$$

implying that

$$x(f_{ij}(y) - f_{ij}(x)) + z(f_{ij}(w) - f_{ij}(z)) \geq 0$$

where  $f_{ij} = \frac{f_i + f_j}{2}$ .

Assume that  $x = z = \frac{y+w}{2}$ . Then

$$\begin{aligned}x(f_{ij}(y) - f_{ij}(x)) + x(f_{ij}(w) - f_{ij}(x)) &\geq 0 \\ f_{ij}(y) - f_{ij}(x) + f_{ij}(w) - f_{ij}(x) &\geq 0 \\ \frac{f_{ij}(y) + f_{ij}(w)}{2} &\geq f_{ij}(x)\end{aligned}$$

so that  $f_{ij}$  is mid-point convex, which for a measurable function implies convexity. Therefore  $f_{ij}$  is differentiable from left and right. We have

$$(y + \epsilon)(f_{ij}(y) - f_{ij}(y + \epsilon)) + (y - \epsilon)(f_{ij}(w) - f_{ij}(y - \epsilon)) \geq 0$$

with equality when  $\epsilon = 0$ . We differentiate with respect to  $\epsilon$  from right.

$$\begin{aligned}(f_{ij}(y) - f_{ij}(y + \epsilon)) + (y + \epsilon)(-f'_{ij+}(y + \epsilon)) \\ - (f_{ij}(w) - f_{ij}(y - \epsilon)) + (y - \epsilon)(f'_{ij-}(y - \epsilon))\end{aligned}$$

which is positive for  $\epsilon = 0$ .

$$\begin{aligned}-y \cdot f'_{ij+}(y) + y \cdot f'_{ij-}(y) &\geq 0 \\ y \cdot f'_{ij-}(y) &\geq y \cdot f'_{ij+}(y).\end{aligned}$$

Since  $f_{ij}$  is convex we have  $f'_{ij-}(y) \leq f'_{ij+}(y)$  which in combination with the previous inequality implies that  $f'_{ij-}(y) = f'_{ij+}(y)$  so that  $f_{ij}$  is differentiable. Since  $f_i = f_{ij} + f_{ik} - f_{jk}$  the function  $f_i$  is also differentiable.

We have

$$\frac{\partial}{\partial q_i} D_F(\rho, \sigma) = p_i f'_i(q_i)$$

and

$$\frac{\partial}{\partial q_i} D_F(\rho, \sigma)|_{\sigma=\rho} = p_i \cdot f'_i(p_i).$$

We have the condition  $\sum q_i = 1$  so using Lagrange multipliers we get that there exist a constant  $c_K$  such that  $p_i \cdot f'_i(p_i) = c_K$ . Hence  $f'_i(p_i) = \frac{c_K}{p_i}$  so that  $f_i(p_i) = c_K \cdot \ln(p_i) + m_i$  for some constant  $m_i$ .

Now we get

$$\begin{aligned} D_F(\rho, \sigma) &= \sum p_i (f_i(q_i) - f_i(p_i)) \\ &= \sum p_i ((c_K \cdot \ln(q_i) + m_i) - (c_K \cdot \ln(p_i) + m_i)) \\ &\quad - c_K \cdot \sum p_i \ln \frac{p_i}{q_i} \\ &= -c_K \cdot D_H(\rho, \sigma). \end{aligned}$$

Therefore there exists an affine function defined on  $K$  such that  $F|_K = -c_K \cdot H|_K + g_K$ . If  $K$  and  $L$  simplices such that  $x \in K \cap L$  then

$$-c_K \cdot H|_K(x) + g_K(x) = -c_L \cdot H|_L(x) + g_L(x)$$

so that

$$(c_L - c_K) \cdot H|_K(x) = g_L(x) - g_K(x).$$

If  $K \cap L$  has dimension greater than zero then the right hand side is affine so the left hand side is affine which is only possible when  $c_K = c_L$ . Therefore we also have  $g_L(x) = g_K(x)$  for all  $x \in K \cap L$ . Therefore the functions  $g_K$  can be extended to a single affine function on the whole of  $\mathcal{S}$ .

Assume that the state  $\sigma$  has two orthogonal decompositions  $\sigma = \sum r_i \rho_i = \sum s_i \sigma_i$ . Since the maximum entropy state  $c$  equals  $\sum \frac{1}{n} \rho_i = \sum \frac{1}{n} \sigma_i$  we have that

$$\begin{aligned} t \cdot \sigma + (1-t) \cdot c &= t \sum r_i \cdot \rho_i + (1-t) \cdot \sum \frac{1}{n} \rho_i \\ &= \sum \left( t \cdot r_i + (1-t) \cdot \frac{1}{n} \right) \rho_i \\ &= \sum \left( t \cdot s_i + (1-t) \cdot \frac{1}{n} \right) \sigma_i. \end{aligned}$$

Therefore

$$\begin{aligned} H(t \cdot \sigma + (1-t) \cdot c) &= - \sum_{i=1}^n \left( t \cdot r_i + (1-t) \cdot \frac{1}{n} \right) \ln \left( t \cdot r_i + (1-t) \cdot \frac{1}{n} \right) \\ &= - \sum_{i=1}^n \left( t \cdot s_i + (1-t) \cdot \frac{1}{n} \right) \ln \left( t \cdot s_i + (1-t) \cdot \frac{1}{n} \right). \end{aligned}$$

Assume  $r_1 \leq r_2 \leq \dots \leq r_n$  and  $s_1 \leq s_2 \leq \dots \leq s_n$ . The analytic continuation of

$t \rightarrow \sum_{i=1}^n (t \cdot r_i + (1-t) \cdot \frac{1}{n}) \ln (t \cdot r_i + (1-t) \cdot \frac{1}{n})$  is defined for values of  $t$  between  $\frac{1}{1-nr_n}$  and  $\frac{1}{1-nr_1}$  and the analytic continuation of  $t \rightarrow \sum_{i=1}^n (t \cdot s_i + (1-t) \cdot \frac{1}{n}) \ln (t \cdot s_i + (1-t) \cdot \frac{1}{n})$  is defined for values of  $t$  between  $\frac{1}{1-ns_n}$  and  $\frac{1}{1-ns_1}$ . Since the functions  $t \rightarrow \sum_{i=1}^n (t \cdot r_i + (1-t) \cdot \frac{1}{n}) \ln (t \cdot r_i + (1-t) \cdot \frac{1}{n})$  and  $t \rightarrow \sum_{i=1}^n (t \cdot s_i + (1-t) \cdot \frac{1}{n}) \ln (t \cdot s_i + (1-t) \cdot \frac{1}{n})$  are identical we most have  $\frac{1}{1-nr_n} = \frac{1}{1-ns_n}$  and  $s_n = r_n$ . Therefore

$$\begin{aligned} &\sum_{i=1}^{n-1} \left( t \cdot r_i + (1-t) \cdot \frac{1}{n} \right) \ln \left( t \cdot r_i + (1-t) \cdot \frac{1}{n} \right) \\ &= \sum_{i=1}^{n-1} \left( t \cdot s_i + (1-t) \cdot \frac{1}{n} \right) \ln \left( t \cdot s_i + (1-t) \cdot \frac{1}{n} \right) \end{aligned}$$

and this argument can be repeated to show that  $r_i = s_i$  that that the two spectra are identical. ■

Combining Theorem 53 with Proposition 30 leads to the following result.

**Corollary 54.** *A weakly state space with a strictly concave entropy function is spectral.*

In [20] it was proved that if a state space is symmetric and spectral then the entropy function is concave.

## VII. CONCLUSION

In this paper we have demonstrated that state spaces are spectral sets if a well-behaved divergence function can be defined. The simple Jordan algebras are symmetric spectral sets and we conjecture that all symmetric spectral sets can be represented on Jordan algebras. A complete classification of spectral sets is highly desirable but does not exist yet.

## ACKNOWLEDGMENT

The author want to thank Prasad Santhanam for inviting me to the Electrical Engineering Department, University of Hawai'i at Mānoa, where some of the ideas presented in this paper were developed. I also want to thank Jan Naudts, Alexander Müller-Hermes, and Chris Perry for useful discussions.

## REFERENCES

- [1] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, vol. 1 of *North-Holland Series in Statistics and Probability*. Amsterdam: North-Holland, 1982.
- [2] P. Harremoës, "Maximum entropy and sufficiency," in *Proceedings MaxEnt2016*, American Institute of Physics (AIP), 2016.
- [3] H. Barnum, J. Barret, M. Krumm, and M. P. Müller, "Entropy, majorization and thermodynamics in general probabilistic theories," arXiv:1508.03107, 2015.
- [4] S. R. Gordon, "Associators in simple algebras," vol. 51, no. 1, pp. 131–141, 1974.
- [5] K. McCrimmon, *A Taste of Jordan Algebras*. Springer, 2004.
- [6] S. L. Adler, *Quaternionic Quantum Mechanics and Quantum Fields*. New York, Oxford: Oxford Univ. Press, 1995.
- [7] E. M. Alfsen and F. W. Shulz, *Geometry of State Spaces of Operator Algebras*. Boston: Birkhäuser, 2003.
- [8] H. Barnum, M. P. Müller, and C. Ududec, "Higher-order interference and single-system postulates characterizing quantum theory," *New Journal of Physics*, vol. 16, no. 12, p. 123029, 2014.
- [9] P. Harremoës, "Divergence and sufficiency for convex optimization," arXiv:1701.01010, 2017.

- [10] A. D. Hendrickson and R. J. Buehler, "Proper scores for probability forecasters," *Ann. Math. Statist.*, vol. 42, pp. 1916–1921, 1971.
- [11] A. Uhlmann, "On the Shannon entropy and related functionals on convex sets," *Reports on Mathematical Physics*, vol. 1, no. 2, pp. 147–159, 1970.
- [12] D. Petz, *Quantum information theory and quantum statistics*. Springer, 2008.
- [13] A. Jenčová and D. Petz, "Sufficiency in quantum statistical inference," *Communications in Mathematical Physics*, vol. 263, no. 1, pp. 259–276, 2006.
- [14] A. Müller-Hermes and D. Reeb, "Monotonicity of the quantum relative entropy under positive maps," *Annales Henri Poincaré*, 2015.
- [15] M. Christandl and A. Müller-Hermes, "Relative entropy bounds on quantum, private and repeater capacities." April, 2016. arXiv:1604.03448, 2016.
- [16] D. Petz, "Sufficiency of channels over von Neumann algebras.," *Quart. J. Math. Oxford*, vol. 39, pp. 97–108., 1988.
- [17] P. Harremoës and N. Tishby, "The information bottleneck revisited or how to choose a good distortion measure," in *Proceedings ISIT 2007, Nice*, pp. 566–571, IEEE Information Theory Society, June 2007.
- [18] J. Jiao, T. C. amd Albert No, K. Venkat, and T. Weissman, "Information measures: the curious case of the binary alphabet," *Trans. Inform. Theory*, vol. 60, pp. 7616–7626, Dec. 2014.
- [19] R. T. Rockafellar, *Convex Analysis*. New Jersey: Princeton Univ. Press, 1970.
- [20] M. Krumm, H. Barnum, J. Barrett, and M. Müller, "Thermodynamics and the structure of quantum theory." arXiv:1608.04461, 2016.